3-23-2017

Application of Enlisted Force Retention Levels and Career Field Stability

Jamie T. Zimmerman

APPLICATION OF ENLISTED FORCE RETENTION LEVELS AND CAREER FIELD STABILITY

THESIS

Jamie T. Zimmermann, Captain, USAF

AFIT-ENS-MS-17-M-167

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-MS-17-M-167

APPLICATION OF ENLISTED FORCE RETENTION LEVELS AND CAREER FIELD STABILITY

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Jamie T. Zimmermann, MS, BS

Captain, USAF

March 2017

AFIT-ENS-MS-17-M-167

APPLICATION OF ENLISTED FORCE RETENTION LEVELS AND CAREER FIELD STABILITY

Jamie T. Zimmermann, MS, BS

Captain, USAF

Committee Membership:

Dr. Raymond R. Hill
Chair

Lt Col Matthew J. Robbins, PhD
Member

AFIT-ENS-MS-17-M-167

Abstract

The Air Force's success is influenced by the manpower it has. The end-strength is Congressionally mandated. Predicting personnel retention is critical to the operations of the military. Over 10 years ago, the Air Force produced enlisted career field sustainment lines based on manning. If a career field was over manned the authorizations were decreased, likewise if a career field was under manned the authorizations were increased. The constant fluctuation of manning caused bathtubs to be created and requirements to go unfilled. Currently, the Air Force produces enlisted career field sustainment lines based on the 5-year historical retention rates. This method produced a more steady state approach, as well as providing a means to adjust the line for other policy actions such as retraining in/out. The need to have a statistically based approach is essential for explaining and defending the creation of the sustainment lines. Many decisions that affect the methods selected for maintaining end-strength are based off the sustainment lines. Data from 2006-2015 was utilized in this research. Logistic regression was used to determine if any significant variables existed, however logistic regression did not provide enough insight into the behavior of the data to be utilized. A survival analysis approach, using retention data, provides a statistically sound methodology to the creation of the sustainment lines. This study produces sustainment lines based on the survival functions for each enlisted career field. It also analyzes the potential of grouping the years of service to manage the career and analyzes retention based on gender and on marital status.

Dedication

This research is dedicated to my supportive parents, my little brother, my wonderful husband and our two children.

Acknowledgments

I would like to thank my advisor, Dr. Raymond R. Hill for his expertise and direction during this research and Lt Col Matthew J. Robbins, PhD, for his leadership and insight into the analysis. I would also like to thank the instructors at the Air Force Institute of Technology for the valuable instruction conducted, providing the resources and tools to complete this research.

Jamie T. Zimmermann

Table of Contents

Page

List of Figures

List of Tables

APPLICATION OF ENLISTED FORCE RETENTION LEVELS AND CAREER FIELD STABILITY

I. Introduction

Problem Background

The United States Air Force (USAF) receives a Congressional end strength cap each fiscal year for the size of the force. End strength is calculated by taking the beginning strength for the fiscal year of personnel, adding the number of personnel gained, and subtracting the number of losses that occurred throughout the fiscal year. These numbers drive the number of people the USAF can recruit and retain each fiscal year. If the end strength number is higher than the current force size then the USAF can recruit more people. However, if the end strength number is lower than the current force size then the USAF must create programs to release people. If the Air Force does not forecast the attrition rate correctly, this could lead to a shortage in manpower or going over the budget constraint. The USAF is given the end strength numbers for a given period in the future so it can plan for the number of recruits.

The Air Force's Enlisted Force is structured based on rank, years of service and Air Force Specialty Codes (AFSCs). Sustainment lines are used by senior leadership to make personnel management decisions. Personnel management decisions include, but are not limited to, retraining, recruiting and separation. Figure 1 provides an example of a sustainment line. The horizontal axis is the years of service (YOS). The vertical axis is the number of personnel. The sustainment line provides a decision maker with a view of what the ideal behavior of the career field would be. The sustainment line is updated every fiscal year. The curve of the line is adjusted for special programs such as retraining in/out and Career Job Reservation (CJR). AFSCs are created and eliminated which contributes to the need to modify manning levels. Approximately, 46 percent of all enlisted losses come from first term attrition (Brauner, Lawson, Mickleson, Adam & Chaiken, 1991).

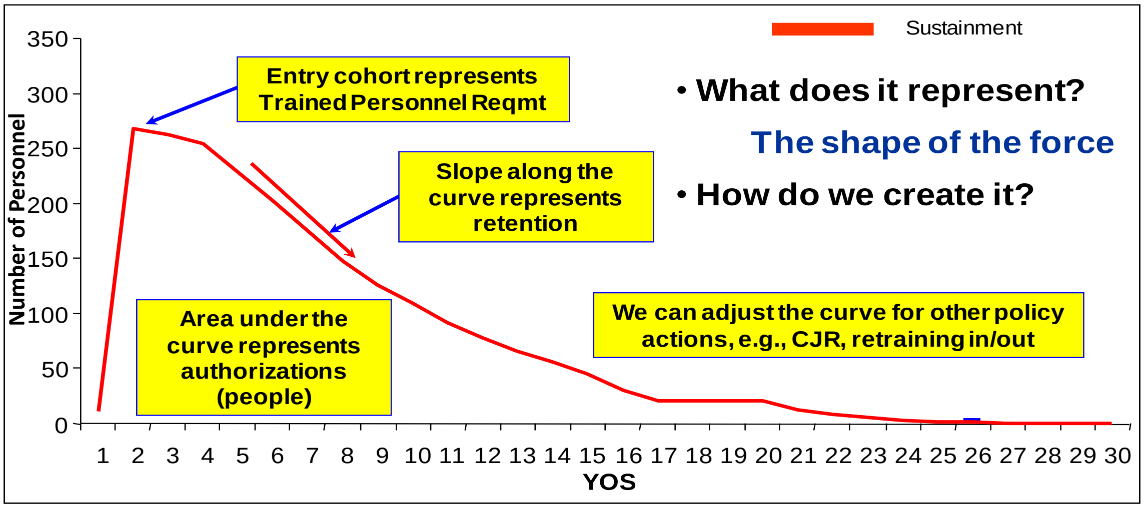


Figure 1. Example of a sustainment line depicting force strength over a 30 year period

Personnel sustainment within any business is critical to their overall success. The Department of Defense is no different. While the size of the force may change, the requirements are continuously increasing. If steps are not taken to predict the behavior of the Air Force, Airmen will continue to work harder to meet mission essential requirements.

Research Scope

The enlisted force is a critical part of daily operations in all the military services, not just the Air Force. Due to the enlisted force's size, a lot of time and money is invested in the training of these Airmen. Retaining the best and brightest in every AFSC should be a priority. This research focuses on modeling the enlisted force retention behavior. Specifically, this research examines a statistically based approach to retention modeling to potentially replace the current sustainment line approach.

Issues, Needs and Limitations

In 2005, the Air Force started to reduce the number of enlisted service members. Sustainment lines are utilized to help determine what AFSCs and year groups are going to see the reduction. Since 2005, force management programs have played an important role in the separations encountered by the enlisted force.

This research utilizes enlisted retention data from 2006-2015. The constant fluctuation in end strength numbers does not appear to be disappearing. Therefore, this data should provide a look into the future as to how the enlisted force structure will behave. The data comes from Headquarters Air Force Directorate of Personnel office (HAF/A1PF) as extracted from the Military Personnel Delivery System (MilPDS). MilPDS is an old system and does not always get updated correctly (Gildea, 2013). For example, if an AFSC is changed there is a possibility that not everyone's record will update accordingly. HAF/A1PF conducted the necessary AFSC conversions before providing the data. Therefore, any assumptions they made about AFSC conversions are inherited into this research.

Thesis Outline

Chapter 2 provides a review of the current literature that exists on the enlisted force personnel management. Chapter 3 explains MilPDS and the data files provided by HAF/A1. Chapter 4 discusses the multivariate analysis and its findings. Chapter 5 explains how the findings were applied to develop an attrition model and assumptions and limitations of the model. Chapter 6 provides a conclusion and recommendations for follow-on research.

II. Literature Review and Background

Chapter Overview

This chapter reviews the relevant past work focused on enlisted military personnel retention. The United States military is a volunteer force. A military member not under an active duty service commitment is eligible to separate from the military service at any point. There are exemptions that allow the member to receive a waiver for their active duty service commitment and separate early. Personnel retention is important when trying to accomplish the mission. Predicting the retention for the force, or some component of the force, for the current year as well as into the future is essential to military success. This chapter presents the various modeling techniques and different analytical methods used in retention modeling.

Modeling Techniques

Personnel programs are developed to help keep the services manned at the Congressional dictated end strength numbers. Models are used to gauge the potential impacts of the programs. These models should provide results leading to insights that can help leadership decide if the policy and/or program should be formalized and implemented.

Models assist analysts in predicting manning levels. Personnel management has been a challenge for decades. Computer-based models were not always available to help solve the problem. As technology advances, military services capability for personnel management also improves. The United States Navy looked at the application of computer-based models for their rotation schedules between sea and shore duty (Jonas, 1979). The actual type of model was not mentioned, only that a computer-based model was used. The model initially focused on enlisted member rotation but was also applied to officer rotation. In 1970, the Naval Personnel and Training Research Laboratory used the rotation model to test and evaluate a proposed personnel policy prior to formalizing and implementing it (Jonas, 1979).

Autoregressive models exploit the time dependency in the observations. Regression models express a relationship among sets of data, autoregressive models express a relationship within some set of data (Brauner et al., 1991). In a first order autoregressive model, each observation correlates with the previous observation. A k^{th} order autoregression models the relationship between observations, k observations apart. An autoregression model could be used to model and then predict retention data collected over time.

A straight line average model uses the average of a specified number of monthly flows to predict the attrition rate (Brauner et al., 1991). If the loss rate is seasonal, a seasonal adjustment factor can be utilized. If the loss model is determined to be seasonal, a seasonal adjustment factor for the given month will be multiplied by the loss rate (Brauner et al., 1991). Seasonality is important when modeling for short-term losses. Such a model could be used to model and predict enlisted applicants using data collected over some number of years. The model might specifically include the high school graduation months as a seasonal factor.

Simulation is used to provide a representation of a real-world system as that system evolves over a period of time. Simulation allows the user to estimate choice probabilities that may arise in behavioral models (Train, 2002). Simulation outputs yield various statistics used to help leadership make decisions. Two key assumptions when using simulation are: the researcher must know how to program the information into the program, and the programmer must understand the assumptions that are built into the program (Train, 2002). A simulation model could be used to represent individuals within units, with those individuals either staying in or leaving the military.

Agent-based simulations can create independent variables/actors, or entities, that have some level of autonomy and internal decision making capabilities. Researchers collect data from the models on how these variables/actors interact with each other in a controlled environment. What-if analyses generally involve changing entity behavior rules. The agent interactions often lead to non-linear outputs which can complicate subsequent statistical analyses. Hill (2006) used agent-based simulation to model pilot retention as a function of three factors believed to affect retention.

Regression analysis is a useful technique for studying military personnel, provided the data are available. Regression analysis helps discover the relationships among variables (Draper & Smith, 2014). In regression analysis, the user hypothesizes a functional relationship among the variables. A response variable from a specified number of factors is plotted in a specified number dimensions. A best fit equation is used to estimate the linear relationship between the variables. The goal is to minimize the estimation error (Montgomery, Peck & Vining, 2012). Regression analysis could be used to examine how well some set of data predicts another set of data. For instance, how well do economic indicators predict Air Force retention (Jantcher, 2016).

Discrete choice models examined with simulation models can be used to forecast personnel actions over time. Such a model describes the decision maker choices from among some set of alternatives (Train, 2002). The discrete set of alternatives is called the choice set. The collective of alternatives must have three characteristics: mutually exclusive, collectively exhaustive and finite. A discrete choice model is normally developed assuming decision maker utility maximizing behavior.

Logit is the most widely used discrete choice model due to the formula for choice probabilities taking a closed form and readily interpretable. A critical part of logit is the assumption that the unobserved factors are uncorrelated over alternatives and all alternatives have the same variance (Train, 2002). While the logit model is convenient, the assumption of independence is not appropriate for all situations.

Neural network techniques are also used in personnel management. The neural networks utilize a processing element, neurons, to develop a solution to a problem. The steps in artificial neural network analysis are mathematically based. The neurons are used to develop a network that allows communication between elements. The network is then trained based on a set of rules. Self-organization or supervised training are used to implement the rules (Wiggins, Engquist & Looper, 1992). Self-organization is when the network does not have a specific goal. Supervised training is when the network does have a specific goal.

Application of Modeling Techniques

Models are only as good as the data that is used in them. Creating models to predict the retention behavior of the unique communities within the Department of Defense can pose challenges that require the user to make key assumptions. The variables used and relationships among the variables will impact the modeling technique utilized.

French (2000) built the Military Personnel Retention Simulator, a computer simulation model that assisted with observing non-linear retention behaviors. Object-oriented programs and multi-agent systems were utilized to provide users with the capability of manipulating the simulated environment and observing the potential retention behaviors. The tool focused on critical commissioned officer specialties and then critical enlisted and warrant officer specialties (French, 2000). The model was not built as a predictive tool but rather as an exploratory tool and thus was not a sole source decision-making tool.

Current Navy manpower models employ a statistical analysis approach (Cashbaugh, 2010). Historical performance data is used to identify correlations between key parameters and retention outcomes. These correlations are exploited to predict future retention outcomes. Cashbaugh (2010) used simulation and optimization to improve Naval personnel analysis and forecasts. He used agent-based simulation to model quality of life factors, social and economic factors and incentives into individual behaviors. The model simulated the Navy's manpower & personnel processes from a bottom-up point of view. Historical survey data was used to correlate attitudes with future behaviors, integrating this attitude data into traditional econometric retention models.

Brauner et al. (1991) developed time series models of the Air Force enlisted force to help predict monthly losses. The Short-Term Aggregate Inventory Projection Model made monthly projections of the aggregate enlisted force for the rest of the current fiscal year. The models were for first-term, second-term and career-term personnel, as well as for extension separation and retirement separation. They used four time series modeling approaches: constant rate, regression, autoregressive and straight line running average. In the constant rate approach, past knowledge is not used (Brauner et al., 1991) and the model takes the form

$$r(t) = c + e(t) \tag{1}$$

where $r(t)$ is the attrition rate at time t,

$e(t)$ is the error at time t, and

c is the mean of the time series.

It was assumed that the error terms ($e(t)$) are independent and normally distributed with a mean of zero and an unknown but constant variance.

Brauner et al. (1991) used regression to see if attrition at time t, $r(t)$, might depend on the airman's salary at time t, $s(t)$,

$$r(t) = (a)*s(t)+c+e(t) \tag{2}$$

where $r(t)$ is the attrition rate at time t,

a is the slope of the linear relationship,

$s(t)$ is the airman's salary at time t,

c is the intercept, and

$e(t)$ is the normally independently distributed error-term with mean zero.

The regression models were examined due to the data exhibiting a dependency on another set of observable variables. When $r(t)$ is plotted against $s(t)$, the relationship looked like a straight line (Brauner et al., 1991).

Brauner et al. (1991) extended (2), adding terms for term of enlistment, number of years to original end of term of service, and grade using

$$r(t) = (a1)*m(t)+a(2)*y(t)+(a3)*g(t)+c+e(t) \qquad (3)$$

where r(t) is the attrition rate at time t,

a1, a2, a3 are the partial regression coefficients,

m(t) is the term of enlistment,

y(t) is the number of years to original end of term of service,

g(t) is the grade of the service member,

c is the intercept, and

e(t) is the normally independently distributed error-term with mean zero.

The model (3) can be described as a multiple regression model since the attrition rate depends on more than one variable.

Orrick (2008) developed a logit model to forecast non-end of active service (NEAS) losses of enlisted Marines by comparing NEAS losses to end of active service (EAS) separation. The logistic models were created using the binary dependent variable denoting the Marine's attrition code. If the Marine retired or died, whether it was combat related or accidental, those observations were removed from the sample. He used Receiver Operating Characteristic (ROC) curves to assess the overall performance of the models. The ROC curves showed the logit models performed well. The models correctly classified NEAS losses with greater than a 76 percent accuracy. However, they misclassified those that were end of active service separations as NEAS losses at a rate below 25 percent (Orrick, 2008). Since the Marine Corps had not used that type of forecasting previously more research was needed before the method could be implemented.

In 2013, the Army was using time series models to forecast active duty army enlisted personnel losses (Whelan, 2013). While time series models are effective in providing predictions, these models do not provide insight as to the underlying causes of the loss behavior. Whelan (2013) used logistic regression to estimate end of term-of-service losses for the army enlisted personnel. This logistic regression analytical technique was appropriate due to the response variable only having two possible outcomes: either they reenlist or they do not. His model estimated the probability that a solider would reenlist given they have 12 months remaining on their current reenlistment contract. Whelan's (2013) model considered a loss to be either the soldier's contract ended or the solider was involuntarily extended by the Army stop-loss policy. Whelan (2013) recommended that his logistic regression model, combined with an attrition model for soldiers who are not in their reenlistment window, could provide a means to make loss estimates for the entire active component enlisted force.

Two recent examples of regression analysis being used for personnel sustainment are Schofield (2015) and Zens (2016). Schofield (2015) examined the retention for non-rated line officers in the United States Air Force. She used logistic regression to determine which factors were significant in predicting retention. Her analysis found gender, number of years as a prior enlisted service member, career field, commissioning year, commissioning source and status of graduation from military education (i.e., Distinguish Graduate) were statistically significant. Schofield built 99 survival models that were comparable to the current model that the Air Force utilizes for manning management. Zens (2016) examined the retention behavior of four Air Force career fields: Acquisition, Non-Rated Operations, Logistics and Support. Her research predicted how many personnel would be serving in that career field for the next 30 years. She used regression and survival analysis. Her analysis found the same factors significant as in Schofield's model.

Wiggins et al. (1992) evaluated artificial neural networks for application to Air Force personnel modeling, specifically reenlistment purposes. Wiggins et al. (1992) also used logit analysis, probit analysis and ordinary least squares for alternate statistical-based comparisons. Logit analysis assumed the cumulative errors followed a logistic distribution. The coefficients in the logit analysis were determined by maximizing the likelihood of observing the actual reenlistment or separation behaviors of the airman in the sample. Probit analysis is similar to logit analysis however in probit analysis there is an assumption of a normal distribution of errors. The output of the linear probability model was interpreted as the probability that the airman will reenlist (Wiggins et al., 1992). The results from the three techniques were compared to three neural network architectures: back propagation, probabilistic network and learning vector quantization, based on the simulation R^2 measure for each model.

In the back propagation architecture, the neural network uses error correction to adapt to the inputs and desired outputs. Wiggins et al. (1992) used the error method that involved minimizing the sum of squared prediction error over all of the training sets. This is similar to linear regression however they applied several nonlinear processing elements to their problem. The use of multiple elements allowed the model to discover the relationship between the inputs and outputs. Since the relationship is not constrained to be linear, back propagation requires more information (i.e., bigger sample size) to find meaningful relationships than regression techniques (Wiggins et al., 1992). The back propagation method required many passes through the data set to complete the training. This is a costly process. One method to improve the performance is to stop training before the network has completely stabilized. This method uses feedback information from the validation sample. This information is not available for making real-time projections over a specific time period or for a specific group of airmen. However, over training is possible and the out-of-sample performance will start to decrease.

The probabilistic neural network forms a separate, non-parametric, probability density function for each of the categories: reenlist or separate. The smoothing factor, which determines the smoothness of the generated probability density function, is the only adjustable factor in the probabilistic neural network. If the smoothing factor is large then the probabilistic density function approximates a multivariate normal distribution centered at the input means of all training in a class (Wiggins et al., 1992).

The learning vector quantization technique solves classification problems. The learning vector quantization technique is similar to the nearest neighbor classifier. In the nearest neighbor classifier, the behaviors of an airman in the validation sample are assumed to be the same as a similar airman from the estimation sample. Similarity can be measured in various ways. Wiggins et al. (1992) used the Euclidean distance between the validation airman's and the estimated airman's input vectors.

Wiggins et al. (1992) also looked at accession and retentions using back propagation networks. These networks, which used only a few neurons, were able to reproduce the estimation sample from fiscal year 1979 through fiscal year 1987 with almost no error. The out-of-sample performance, using the same networks, was poor.

During training, Wiggins et al. (1992) used a heuristic with the time series data to stop the process before over-fitting occurred.

Wiggins et al. (1992) used three heuristics: 79 hold-out, 88 hold-out, and inflections. The 79 hold-out heuristic selected the amount of training that produced the best out-of-sample performance on the January 1979 through September 1979 sample. The 88 hold-out heuristic was similar except the time frame was October 1987 through September 1988. The inflections heuristic did not utilize information outside of the training sample. Training was stopped at the second negative to positive inflection in the root mean square error of the in-sample training path. For the reenlist portion of their research, they used a sigmoid function. For the retention portion, they used a hyperbolic tangent function. The time series data helped prove that the networks with hyperbolic tangents were consistently training to obtain similar results with similar training epochs (Wiggins et al, 1992). The hyperbolic function required a linear transformation of the output variables. However, this did not affect the simulation R^2 value. Two additional modeling techniques were utilized: Ordinary Least Squares and Generalized Least Squares. All modeling techniques performed well. The inflection technique continuously stopped the training too early, thus leading to the inflection technique performing the worst among the neural network techniques.

Overall, neural networks have the ability to improve performance on the models. Overfitting is an area of concern when conducting research on personnel systems. The heuristics that Wiggins et al. (1992) utilized proved effective in preventing the network from over generalizing.

II. Data Source

The data utilized in this research originates from the Military Personnel Delivery System (MilPDS). MilPDS is a database that houses the primary records for personnel data and actions throughout an Airmen's career. Accessions, reenlistments and separations are recorded in MilPDS (Gildea, 2013). MilPDS is considered a precise and up-to-date way to keep track of Airmen's personnel careers. While the system has undergone some critical updates, human error does occur. Thus Airmen are encouraged to check their records for accuracy. One error that may occur is the creating, deleting or combining of Air Force Specialty Codes. While every effort is made to conduct a smooth transition, there are records that do not make it through the transition and must be manually corrected.

Air Force Manpower, Personnel and Services (AF/A1PF) supplied the data used in this research. The data was provided in the form of SAS sheets, and syntax was written to perform the analysis. The data covered all Air Force enlisted members, both retained and not retained. The inventory files contain various information about a service member to include gender, race, marital status, duty location, previous AFSCs and number of dependents. Loss files are created by a "if there and not here" look. If the member's social security number was listed in one month's data set however was not listed in the next month's, the service member is considered a loss and is included in the loss files for that month.

There is a multitude of enlisted career fields. The four selected to examine utilizing logistic regression were selected on a random basis. The selection attempted to capture four diverse career fields. The databases contained information for each of the variables analyzed. At least one career field had to be receiving a Selective Re-enlistment Bonus (SRB) to ensure that portion of the analysis was possible. Career fields with shreds and retrain-in only career fields were not considered during the selection due to the complexity of their structure behavior.

For survival analysis, a survival curve was produced for each Air Force Specialty Code (AFSC). The records that contained errors, such as “BLANK” in the Control Air Force Specialty Code Trained Personnel Resources (CTPR) variable spot, were grouped and a curve was created for those records as well. The data is used in this analysis is the same data that A1PF will have access to on a reoccurring basis. The survival curves are dependent on the accuracy and reliability of the data source.

IV. Analysis

Logistic regression was used to determine which factors were significant in predicting Air Force enlisted retention. The response variable was binary. The loss record was assigned a zero and the retrained record was assigned a one. The variables examined were grade, sex, race, marital status, dependents and years of service. The number of dependents variable (DEPNTONR) was modified to a binary variable (DEPENDS). If the service member's record indicated at least one dependent, the DEPENDS variable was assigned a one else a zero. The data was examined at an aggregate level, separated by year, separated by AFSC and separated by Selective Reenlistment Bonus (SRB) zone for one AFSC. The four AFSCs that were selected to examine individually were Airfield Management (1C7X1); Operations Intel (1N0X1); Survival; Evasion; Resistance and Escape (1T0X1); and Mental Health Services (4C0X1). Airfield Management was selected to examine based on SRB zone.

Data from 2006 to 2015 was provided by month, and SAS was utilized to combine the inventory data into years. The same process was applied to the loss files. Once the files were combined, duplicate records were removed. The first iteration of logistic regression was completed at an aggregate level using SAS. The second iteration of logistic regression was completed by fiscal year, again using SAS. Appendix A contains the SAS code for the compilation of data and the logistic regressions. The data was then imported into JMP to conduct logistic regression on the four selected AFSCs and the one AFSC by SRB zone. JMP is a product of SAS. Table 1 shows the p-values of the six variables at an aggregate level for 2006-2015 data. All six variables are significant at this level. This model produced an Rsquare of 0.06139 with an adjusted

Rsquare of 0.061321. Based on the adjusted Rsquare this model is not a good fit to predict enlisted retention.

Table 1. Analysis of Effects, 2006-2015, Data Summary of P-Values

Grade	Sex	Race	Marital Status	Dependents	Years of Service
<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table 2 breaks the data into year groups. In all the years except 2006, 2007, 2010 and 2011, all six factors are found to be significant. Once the data was broken out by years, inconsistency with significant variables was present. Marital Status was the factor that changed in significance from, year-to-year. Due to this event, it is not certain if marital status is a significant factor.

Table 2. Analysis of Effects by Year, Summary of P-Values

Year	Grade	Sex	Race	Marital Status	Dependents	Years of Service
2006	<0.0001	<0.0001	<0.0001	0.2217	<0.0001	<0.0001
2007	<0.0001	<0.0001	<0.0001	0.0819	<0.0001	<0.0001
2008	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
2009	<0.0001	<0.0001	<0.0001	0.0052	<0.0001	<0.0001
2010	<0.0001	<0.0001	<0.0001	0.1782	<0.0001	<0.0001
2011	<0.0001	<0.0001	<0.0001	0.2340	<0.0001	<0.0001

2012	<0.0001	<0.0001	<0.0001	0.0138	<0.0001	<0.0001
2013	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
2014	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
2015	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Table 3 breaks the data into year groups for the four selected AFSCs. The results do not provide useful insight into the retention of enlisted personnel. Years of Service and Grade are correlated. A service member has to have a minimum time in grade which is related to years of service before achieving the next grade. The decomposing of data further showed more inconsistencies of identifying significant factors among AFSCs and years. For example, 1N0X1 Grade was a significant factor in each year except 2008 and Sex was only significant in 2006, 2007 and 2008. These inconsistencies cause the models to be inadequate.

Table 3. Analysis of Effects by AFSC, Summary of P-Values

Year	AFSC	Grade	Sex	Race	Marital Status	Dependents	Years of Service
2006	1C7X1	0.0001	0.3238	0.7803	0.3677	0.2448	<0.0001
	1N0X1	0.0002	0.0002	0.9315	0.3032	0.8454	<0.0001
	1T0X1	0.1402	0.1018	0.5248	0.9742	0.2779	0.0040
	4C0X1	0.1074	0.2184	0.7242	0.0896	0.7655	0.0362
2007	1C7X1	0.0046	0.9469	0.5111	0.0512	0.1710	0.0009
	1N0X1	0.0184	<0.0001	0.5258	0.7596	0.7867	<0.0001

	1T0X1	0.0037	0.0293	0.9273	0.5317	0.8789	<0.0001
	4C0X1	0.1717	0.1479	0.9224	0.8769	0.1206	0.0213
2008	1C7X1	0.0195	0.0016	0.5153	0.0301	0.0257	0.0004
	1N0X1	0.2260	<0.0001	0.0958	<0.0001	<0.0001	0.0086
	1T0X1	0.2820	0.3933	0.0845	0.0197	0.2295	0.8965
	4C0X1	0.0226	0.1004	0.0833	0.0548	<0.0001	0.0045
2009	1C7X1	<0.0001	0.0431	0.4965	0.9654	0.7174	<0.0001
	1N0X1	0.0002	0.0642	0.6306	0.1405	0.2892	<0.0001
	1T0X1	0.5467	0.3761	0.9217	0.8598	0.1423	0.0095
	4C0X1	0.0015	0.3403	0.3063	0.4259	0.5381	<0.0001
2010	1C7X1	0.0005	0.0016	0.1188	0.8198	0.0837	<0.0001
	1N0X1	0.0053	0.1605	0.9751	0.7594	0.7890	<0.0001
	1T0X1	0.0078	0.3971	0.8658	0.1502	0.1635	<0.0001
	4C0X1	0.0003	0.9477	0.7745	0.1109	0.0387	<0.0001
2011	1C7X1	0.0001	0.7932	0.8749	0.1105	0.0801	<0.0001
	1N0X1	<0.0001	0.5998	0.3982	0.0013	0.0002	<0.0001
	1T0X1	<0.0001	0.2120	0.6020	0.0985	0.5137	<0.0001
	4C0X1	<0.0001	0.9629	0.8051	0.7670	0.1910	<0.0001
2012	1C7X1	<0.0001	0.3384	0.1329	0.6877	0.2137	<0.0001
	1N0X1	0.0096	0.3251	0.1834	0.4189	0.0344	0.0010
	1T0X1	0.0117	0.4743	0.9027	0.0970	0.0662	0.0017
	4C0X1	<0.0001	0.2106	0.9387	0.3187	0.3533	<0.0001

2013	1C7X1	<0.0001	0.1112	0.9210	0.1490	0.3185	0.0001
	1N0X1	<0.0001	0.9954	0.0233	0.8897	0.0001	<0.0001
	1T0X1	0.0313	0.4683	0.4092	0.4180	0.1591	0.3195
	4C0X1	0.0798	0.2162	0.6163	0.8169	0.1493	0.0801
2014	1C7X1	<0.0001	0.2241	0.5062	0.5217	0.0655	0.0001
	1N0X1	<0.0001	0.8921	0.7655	0.2883	0.0303	0.0001
	1T0X1	0.0006	0.6108	0.8532	0.8299	0.8881	0.0004
	4C0X1	0.0004	0.3335	0.8735	0.3018	0.5144	0.0025
2015	1C7X1	<0.0001	0.3851	0.3583	0.9756	0.3505	<0.0001
	1N0X1	0.0015	0.1340	0.2983	0.0126	0.2293	0.0163
	1T0X1	0.2391	0.4009	0.3040	0.0949	0.2150	0.0951
	4C0X1	0.0207	0.2544	0.2051	0.3687	0.9830	0.0091

Finally, Table 4 breaks the 1C7X1 AFSC into SRB zones. The zones are identical to the way the Air Force pays SRBs. If the p-values are not present, then the data for 1C7X1 SRB zone for that year was not available. The inconsistencies of significant factors are present when looking at an AFSC by zone and year. It is not expected that the zones would agree on significate factors. However, if the information were going to be insightful, one would expected there to be consistencies among the same zone but in different years. This did not occur, as shown in Table 4. The yellow boxes indicate the factor is significant. The inconsistencies lead the models to be inadequate.

Table 4. 1C7X1 Analysis of Effects by Zone Summary of P-Values

Year	Zone	Grade	Sex	Race	Marital Status	Dependents	Years of Service
2006	A	0.0910	0.1530	0.6214	0.2644	0.8278	0.2024
	B	0.0003	0.8867	0.9400	0.4568	0.3632	0.8736
	C	0.0608	0.6828	0.0364	0.0059	0.0177	0.8953
	E	-	-	-	-	-	-
2007	A	0.0806	0.8479	0.1123	0.2341	0.1607	0.0001
	B	0.0069	0.1758	0.2187	0.0753	0.1672	0.4487
	C	0.7257	0.9104	0.7430	0.9066	0.3018	0.1857
	E	-	-	-	-	-	-
2008	A	0.6555	0.3939	0.2895	0.7088	0.7155	0.0092
	B	0.0002	0.0078	0.6670	0.0575	0.0276	0.2411
	C	0.6346	0.0093	0.5751	0.3892	0.1089	0.6517
	E	-	-	-	-	-	-
2009	A	<0.0001	0.9577	0.9298	0.4584	0.0947	0.0002
	B	0.0224	0.0042	0.1900	0.0397	0.0723	0.2844
	C	0.0425	0.3084	0.4533	0.6928	0.3569	0.5109
	E	1.0000	1.0000	0.8260	0.9996	1.0000	0.4613
2010	A	0.0114	0.0189	0.8919	0.5790	0.2357	0.0171
	B	0.0161	0.0011	0.4172	0.9815	0.6440	0.0170

	C	0.8492	0.4491	0.6820	0.0283	0.0063	0.7901
	E	-	-	-	-	-	-
2011	A	0.0146	0.8754	0.6781	0.1569	0.2088	0.0080
	B	0.0191	0.6364	0.7877	0.4652	0.1051	0.0068
	C	<0.0001	0.0070	0.6054	0.0030	0.0959	0.0008
	E	-	-	-	-	-	-
2012	A	0.0006	0.7612	0.3396	0.6433	0.4701	0.0008
	B	0.2649	0.6195	0.3917	0.0262	0.3837	0.8499
	C	<0.0001	0.3344	0.0015	0.2547	0.0868	0.6597
	E	-	-	-	-	-	-
2013	A	0.0011	0.0109	0.9185	0.8012	0.7803	0.0085
	B	0.0997	0.3868	0.9448	0.6427	0.4586	0.1663
	C	0.1831	0.6698	0.5449	0.3101	0.3122	0.7346
	E	-	-	-	-	-	-
2014	A	0.0006	0.9454	0.2807	0.4027	0.0113	0.0003
	B	<0.0001	0.1674	0.2638	0.2879	0.9775	0.4745
	C	0.1311	0.0047	0.5821	0.2731	0.9851	0.7479
	E	0.6002	0.1154	1.0000	1.0000	0.9999	0.2654
2015	A	0.0167	0.4400	0.1402	0.4575	0.2000	0.0034
	B	0.0093	0.0629	0.3005	0.8083	0.8322	0.0424
	C	0.1056	0.4864	0.9673	0.9379	0.0580	0.3568
	E	-	-	-	-	-	-

As the data was decomposed into more levels, the factors became less significant. This can create more challenges if an individual were to try to manage a career field at this level of decomposition.

Building a logistic regression model on retention poses problems. The Air Force enlisted data is quite noisy. Therefore, retention trends by career field are examined using survival analysis.

Survival analysis models factors that influence the time to an event. The event in this research is the retention of a service member. Therefore, retention data was used for the survival analysis. One challenge is the use of censored data. While the survival function is providing information about the retained service members, we have limited insight as to when that service member will leave the Air Force. In the data, the retain variable was binary (0=loss and 1=retained). SAS was utilized for the survival analysis; SAS code is in Appendix B. The function proc lifetest is a nonparametric estimate of the survivor function using either the Kaplan-Meier method or the actuarial method. Appendix D contains all the survival function graphs produced from the SAS code.

The hazard function was computed in SAS using the PROC PHREG function. The code for Cox regression is in Appendix B. The hazard rate represents the instantaneous probability that the service member separates at a certain years of service, given that the service member has lasted up to the years of service. Logistic regression estimates the odds ratios where Cox regression estimates the hazard ratio. Cox regression can handle both discrete and continuous measures of event times (Sainani, n.d).

After calculating the survival and hazard function, the AFSC was examined further by breaking the years of service into bins. The first iteration contained two bins, years 0-20 and years 20+. The second iteration contained four bins, years 0-4, years 5-10, years 11-20 and years 20+. Logistic regression, survival analysis and hazard functions were created for both iterations using SAS. The SAS code can be viewed in Appendix C. The purpose of these two iterations was to see if any insight could be gathered from breaking the career fields down by years of service. The act of breaking down the AFSCs by years of service provided the decision maker with a zoomed view of the group. There are drops in years of service that may not have been overlooked when the AFSC is analyzed as a whole.

V. Application

The survival graphs provide A1PF with the retention behavior for enlisted AFSCs. The AFSCs that are retraining-in only, have extended school houses, and special duties should be analyzed on an individual basis. An example of a retraining-in only AFSC is 7S0X1, Special Investigations. An example of an AFSC with an extended school house is 1A8X1M, Airborne Crypto Language Analyst—Pashto. An example of a special duty AFSC is 9E000, Command Chief Master Sergeant. The shapes of the survival graphs were compared to the current sustainment lines and proved to be accurate. The height of the sustainment lines will vary from year to year depending on the number of authorizations for each AFSC.

Figure 2 shows the AFSC's 1A0X1, In-Flight Refueling, behavior utilizing the SAS code for survival analysis. Figure 3 shows the same AFSC however the red sustainment line was developed utilizing A1PFs current methodology. The red sustainment line has been smoothed. The purple line is the grade authorizations however this information was not utilized in this research. The blue bars is the permanent party inventory and the yellow bars represent the number of people in student, transient or prisoner status. The overall shape of the line between the two figures is similar.

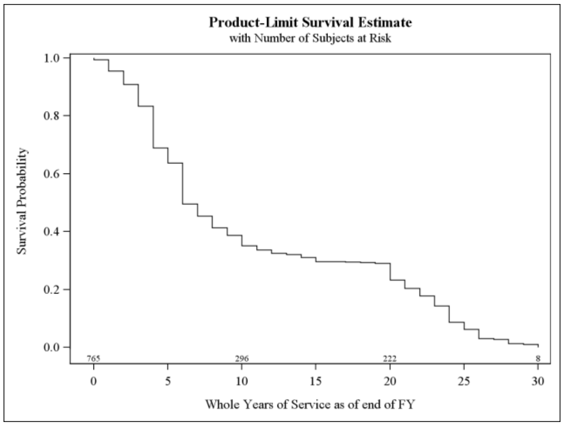


Figure 2. Example of 1A0X1 sustainment line using survival analysis

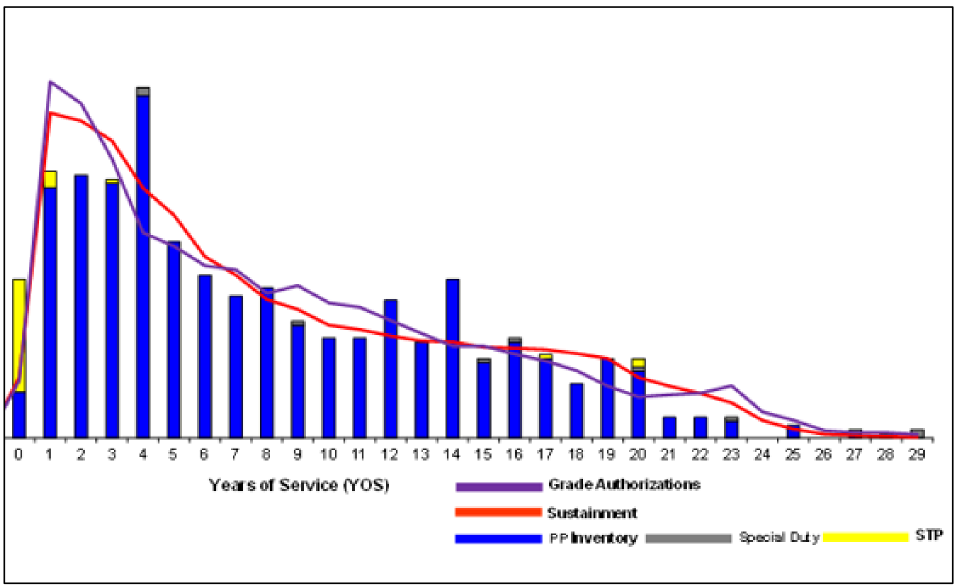


Figure 3. Example of 1A0X1 current sustainment line

Figure 4 is the sustainment line using survival analysis for a retraining-in only AFSC, 7S0X1, Special Investigations. Figure 5 is the sustainment line using the current sustainment methodology for the same AFSC. The sustainment line with the current methodology is in more of a bell shape due to the nature of the AFSC. The line created utilizing survival analysis appears to look more traditional. This AFSC is not a traditional AFSC and therefore the approach to create the sustainment may potentially not be similar to other AFSCs.

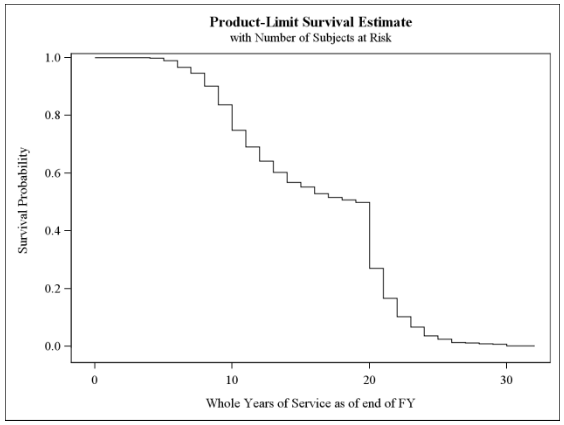


Figure 4. Example of 7S0X1 sustainment line using survival analysis

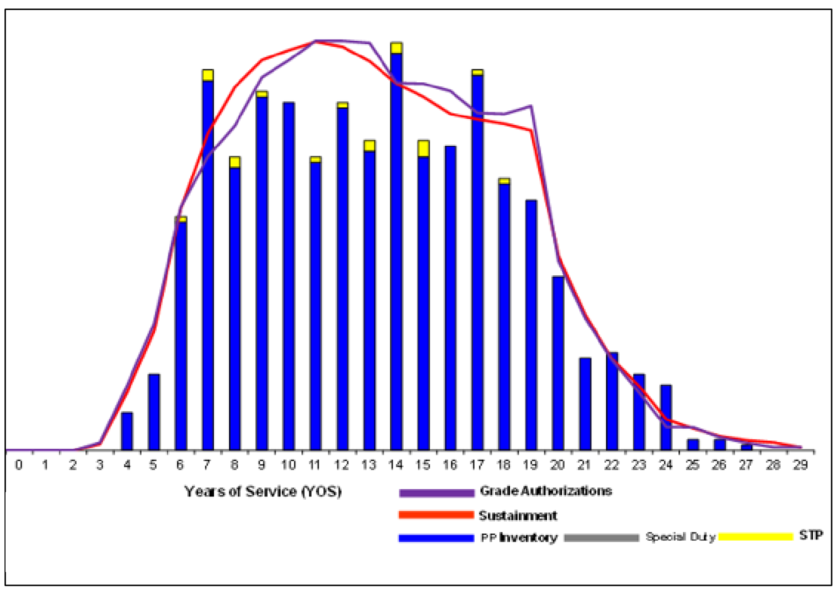


Figure 5. Example of 7S0X1 current sustainment line

Figure 6 is the sustainment line using survival analysis for an AFSC with a shred, 1A8X1M, Airborne Crypto Language Analyst--Pashto. Figure 7 is the sustainment line using the current sustainment methodology for the same AFSC. The graph utilizing survival analysis shows a greater drop in inventory at year 6 than the current sustainment line methodology. Career Field Managers would be contacted for their subject matter expert opinion on the behavior of the shred.

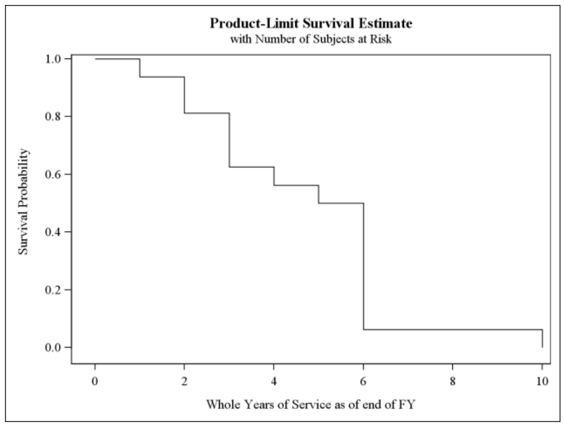


Figure 6. Example of 1A8X1M sustainment line using survival analysis

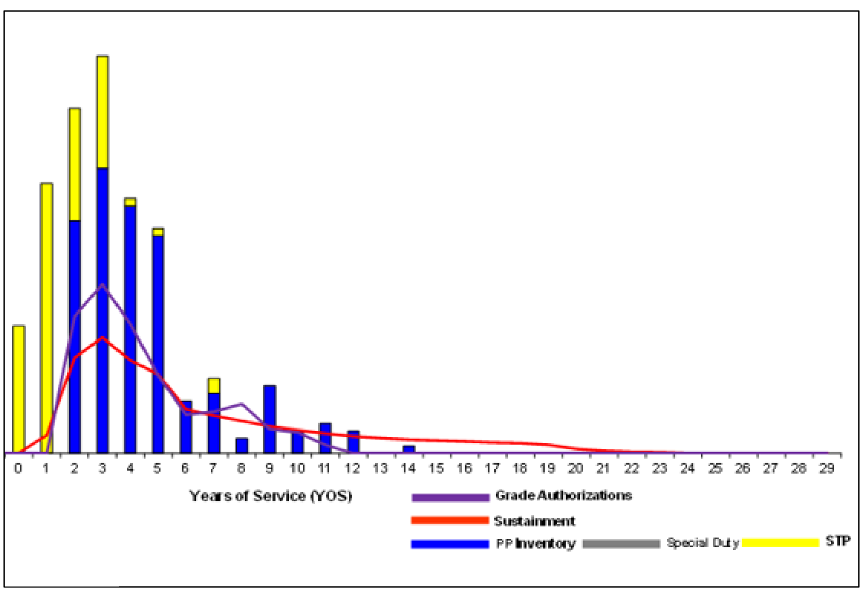


Figure 7. Example of 1A8X1M current sustainment line

Figure 8 is the sustainment line using survival analysis for a special duty AFSC, 9E000, Command Chief Master Sergeant. Figure 9 is the sustainment line using the current sustainment methodology for the same AFSC. Figure 9 shows a peak at year 27 however figure 8 is showing a constant decreasing curve. In this case, the survival analysis was not the optimal method to depict the behavior of this special duty AFSC.

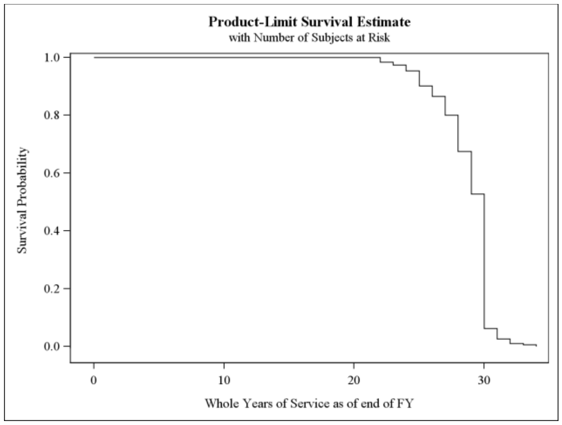


Figure 8. Example of 9E000 sustainment line using survival analysis

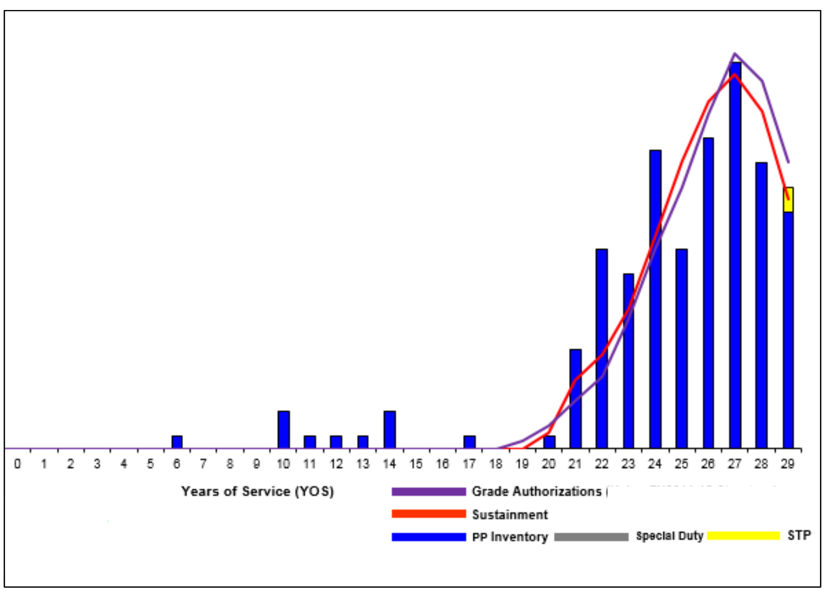


Figure 9. Example of 9E000 current sustainment line

Enlisted Career Field Managers can provide more information that might influence the overall behavior of the career field. While the survival graphs are useful, the Career Field Managers should be used when outputs look suspicious and/or major events to the career field are going to occur. Some AFSCs break out into shreds and the retention behavior of service members to and from the shreds may vary between all AFSC shreds.

Using the strata function in conjunction with the lifetest, graphs can be created based on specified variables. While Air Force retention as a whole is important, diversity is something that leadership strives to accomplish. Based on figure 10, females show a lower retention rate around year 4. While many hypotheses can be created as to why this may accord, this graph can show leadership whether gender diversity is something that should be analyzed.

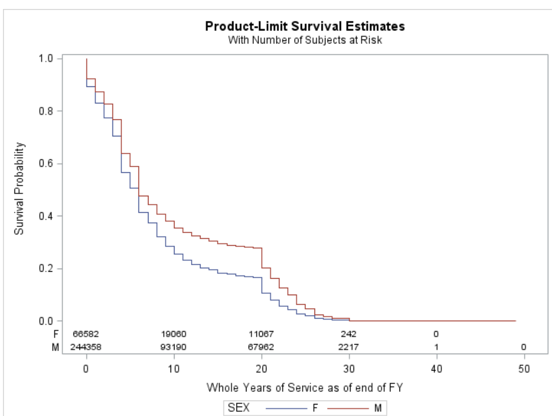


Figure 10. Survival Analysis Graph by Gender

Another variable that was broken out is marital status, Figure 11, A: Annulled, D: Divorced, L: Legal Separation, M: Married, S: Single and W: Widowed. Marital status can affect a personal mentally and physically. Understand the retention patterns of the various marital statuses can provide valuable insight to leadership. For example, the retention rate for single service members appears to be lower than other marital statuses. Again, hypotheses may be tested to try and figure out the cause of this observation however this graph shows leadership that something is going on with the single service members. Leadership may want to create programs specifically for single service members to attract them and retain them.

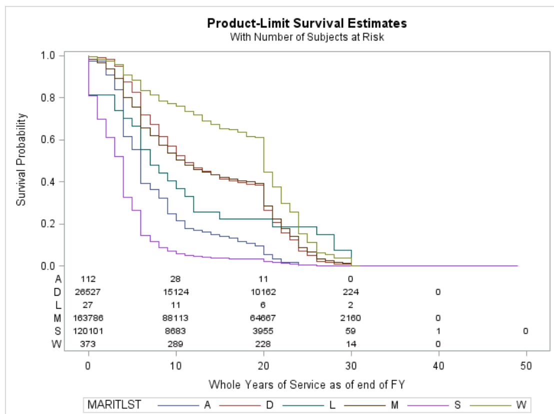


Figure 11. Survival Analysis Graph by Marital Status

Although breaking out the year groups into bins did not provide useful insight into the career field behavior, survival analysis graphs are available on the supplement compact disc. Additional AFSC survival analysis graphs and hazard function graphs are available on the supplement compact disc.

VI. Conclusion

The Air Force enlisted AFSCs are not all the same. There are AFSCs that are retraining-in only, have extended school houses, receive SRBs, special duties and that allow retraining-in and retraining-out. This causes many challenges to arise when analyzing the retention behavior and trying to create a blanket methodology for all AFSCs. If the Milpds records are complete, SAS provides a useful way to calculate and create the survival graphs for every AFSC. While the special AFSCs may require additional attention, the decision maker will have a starting point to work with.

Being able to predict what a career field will look like in future years is essential to the success of the career field. The analyst needs to be able to provide a sound methodology if questioned about their methods in creating the sustainment lines. This research provides statistical methodology. The survival lines were similar to the current sustainment lines that A1PF calculates and utilizes.

It is recommended that A1PF utilize SAS and the proc lifetest function to create further sustainment lines. Any AFSCs that have exceptions to the normal operation should be analyzed on a base by case basis, as A1PF currently does. The strata function provides more insight into the AFSC depending on the variable selected for the analysis. This may prove to be useful when discussing the need for diversity and trying to create programs that are targeted to certain groups of people. While A1PF does not specialize in diversity or the creation of airman programs, they have opportunities to be involved in discussions where these graphs may prove to be valuable.

This research looked at AFSCs as a whole and broke them into bins. It also looked at retention by gender and retention by marital status. Future research should include deployment information. There are AFSCs that are deployed more often than others and this may influence the retention of that service member. Another factor to look at is the SRB effects on retention. This bonus is an incentive for service member to reenlist, when eligible, and receive a bonus at the same time. Normally, these bonuses are offered to AFSCs that have a retention problem in a certain zone.

A1PF does not necessarily manage the career fields by 3-letter however rolling up the AFSC to 3 letter may provide insightful information about the retention behavior. The one challenge with this method will be the payment of SRBs are conducted on an AFSC basis, there are shreds and controlling the retraining ins and outs of various AFSCs. Rolling up to a 3 letter might actually cause the decision maker to lose fidelity on current and arising problems.

The blended retirement system will eventually change the retention behaviors. In the past, if a service member completed 10 years of service it was assumed that they will continue to 20 years of service for the retirement. If that service member separated from the service before 20 years of service, he/she would not receive any type of retirement. The new blended retirement system eliminates that problem so assuming a service will stay in to 20 years of service purely due to them reaching 10 years of service will not be a valid assumption.

Data mining is a useful tool however without including the political, social and economic aspects, data mining will not provide the whole picture. The economic status will affect each AFSC differently. There are AFSCs that translate directly to outside careers and others that are special to the military. Predicting retention is challenging and there are many outside factors that influence an individual’s decision to separate or retain.

Each AFSC has different requirements as far as deployment, training, work schedules and overall demand. The survival function in SAS provides an overall methodology without creating something special for each AFSC and editing it each year when the sustainment lines are updated.

Appendix A. SAS Code for Logistic Regression 2006-2015 Data

```
*Combine 2006 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2006";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200601"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200602"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200603"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200604"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200605"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200606"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200607"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200608"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200609"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200610"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200611"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200612";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine 2006 Enlloss2006;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2006";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200601"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200602"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200603"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200604"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200605"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200606"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200607"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200608"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200609"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200610"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200611"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200612";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2006 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2006";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2006"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2006";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2006";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2006";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
```

```
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput06.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2006";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;

*Combine 2007 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2007";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200701"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200702"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200703"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200704"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200705"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200706"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200707"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200708"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200709"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200710"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200711"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200712";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine 2007 Enlloss2007;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2007";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200701"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200702"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200703"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200704"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200705"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200706"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200707"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200708"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200709"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200710"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200711"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200712";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2007 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2007";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2007"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2007";
      by SSAN;
```

```
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2007";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2007";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput07.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2007";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;

*Combine 2008 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2008";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200801"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200802"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200803"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200804"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200805"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200806"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200807"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200808"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200809"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200810"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200811"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200812";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine Enlloss2008;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2008";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200801"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200802"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200803"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200804"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200805"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200806"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200807"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200808"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200809"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200810"
```

```
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200811"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200812";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2008 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2008";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2008"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2008";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2008";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2008";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput08.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2008";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Combine 2009 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2009";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200901"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200902"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200903"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200904"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200905"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200906"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200907"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200908"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200909"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200910"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200911"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv200912";
      by SSAN;
      if last.SSAN;
      retain = 1;
```

```
run;
*Combine  Enlloss2009;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2009";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200901"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200902"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200903"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200904"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200905"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200906"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200907"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200908"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200909"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200910"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200911"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss200912";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2009 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2009";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2009"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2009";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2009";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2009";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput09.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2009";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;

*Combine 2010 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2010";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201001"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201002"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201003"
```

```
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201004"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201005"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201006"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201007"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201008"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201009"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201010"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201011"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201012";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine  Enlloss2010;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2010";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201001"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201002"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201003"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201004"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201005"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201006"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201007"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201008"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201009"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201010"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201011"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201012";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2010 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2010";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2010"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2010";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2010";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2010";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput10.doc";
```

```
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2010";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;

*Combine 2011 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2011";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201101"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201102"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201103"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201105"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201106"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201107"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201108"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201109"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201110"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201111"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201112";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine  Enlloss2011;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2011";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201101"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201102"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201103"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201104"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201105"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201106"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201107"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201108"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201109"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201110"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201111"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201112";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2011 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2011";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2011"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2011";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2011";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2011";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
```

```
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput11.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2011";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;



*Combine 2012 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2012";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201201"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201202"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201203"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201204"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201205"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201206"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201207"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201208"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201209"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201210"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201211"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201212";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine  Enlloss2012;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2012";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201201"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201202"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201203"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201204"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201205"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201206"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201207"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201208"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201209"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201210"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201211"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201212";
by SSAN;
if last.SSAN;
retain = 0;
```

```
run;
*Combine 2012 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2012";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2012"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2012";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2012";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2012";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput12.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2012";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Combine 2013 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2013";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201301"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201302"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201303"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201304"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201305"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201306"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201307"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201308"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201309"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201310"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201311"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201312";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine  Enlloss2013;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2013";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201301"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201302"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201303"
```

```
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201304"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201305"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201306"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201307"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201308"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201309"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201310"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201311"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201312";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2013 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2013";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2013"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2013";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2013";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2013";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput13.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2013";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Combine 2014 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2014";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201401"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201402"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201403"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201404"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201405"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201406"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201407"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201408"
```

```
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201409"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201410"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201411"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201412";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine  Enlloss2014;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2014";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201401"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201402"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201403"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201404"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201405"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201406"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201407"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201408"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201409"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201410"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201411"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201412";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2014 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2014";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2014"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2014";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2014";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2014";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput14.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2014";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;
```

```
*Combine 2015 Enlinv Files;
data "I:\setup\Desktop\Thesis\SASCombine\Enlinv2015";
set "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201501"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201502"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201503"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201504"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201505"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201506"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201507"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201508"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201509"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201510"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201511"
      "I:\setup\Desktop\Thesis\ENLIV_DATA_UnZipped\enlinv201512";
      by SSAN;
      if last.SSAN;
      retain = 1;
run;
*Combine  Enlloss2015;
data "I:\setup\Desktop\Thesis\SASCombine\Enlloss2015";
set "I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201501"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201502"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201503"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201504"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201505"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201506"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201507"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201508"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201509"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201510"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201511"
"I:\setup\Desktop\Thesis\ENLLOSS_DATA_UnZipped\enlloss201512";
by SSAN;
if last.SSAN;
retain = 0;
run;
*Combine 2015 Enlinv and Enlloss files;
data "I:\setup\Desktop\Thesis\SASCombine\Final2015";
set "I:\setup\Desktop\Thesis\SASCombine\Enlinv2015"
      "I:\setup\Desktop\Thesis\SASCombine\Enlloss2015";
      by SSAN;
      if last.SSAN;
run;
DATA "I:\setup\Desktop\Thesis\SASCombine\Final2015";
    SET "I:\setup\Desktop\Thesis\SASCombine\Final2015";
      If DEPNTONR = '' then DEPENDS = 0;
      If DEPNTONR = '01' then DEPENDS = 1;
      If DEPNTONR = '02' then DEPENDS = 1;
      If DEPNTONR = '03' then DEPENDS = 1;
      If DEPNTONR = '04' then DEPENDS = 1;
      If DEPNTONR = '05' then DEPENDS = 1;
      If DEPNTONR = '06' then DEPENDS = 1;
      If DEPNTONR = '07' then DEPENDS = 1;
```

```
      If DEPNTONR = '08' then DEPENDS = 1;
      If DEPNTONR = '09' then DEPENDS = 1;
      If DEPNTONR ge '10' then DEPENDS = 1;
RUN;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput15.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\Final2015";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Combine All FILES*;
data "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
set "I:\setup\Desktop\Thesis\SASCombine\Final2006"
      "I:\setup\Desktop\Thesis\SASCombine\Final2007"
      "I:\setup\Desktop\Thesis\SASCombine\Final2008"
      "I:\setup\Desktop\Thesis\SASCombine\Final2009"
      "I:\setup\Desktop\Thesis\SASCombine\Final2010"
      "I:\setup\Desktop\Thesis\SASCombine\Final2011"
      "I:\setup\Desktop\Thesis\SASCombine\Final2012"
      "I:\setup\Desktop\Thesis\SASCombine\Final2013"
      "I:\setup\Desktop\Thesis\SASCombine\Final2014"
      "I:\setup\Desktop\Thesis\SASCombine\Final2015";
      by SSAN;
      if last.SSAN;
run;

ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutputALL.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;
```

Appendix B. SAS Code for Survival and Hazard Analysis 2006-2015 Data

```
*Sort the data by CTPR;
proc sort data = "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
by CTPR;
run;

ods graphics on;
*Survival function;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"(where=(retain=0)
) plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;


*Survival Function Export only Graphs;
ods graphics on;
ods select survivalplot(persist);
ODS RTF FILE="I:\setup\Desktop\Thesis\SASlifetestplots.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"(where=(retain=0)
) plots=survival(atrisk);
by CTPR;
Title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;

*Cox Regression Model;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput4Try.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;
```

```
*Survival function Retention by Gender;
ods graphics on;
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"(where=(retain=0)
) plots=survival(atrisk);
*by CTPR;
Strata Sex;
time YOS_EFY*retain(1);
run;
ods graphics off;

*Survival function Retentio by Marital Status;
ods graphics on;
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"(where=(retain=0)
) plots=survival(atrisk);
*by CTPR;
Strata MARITLST;
time YOS_EFY*retain(1);
run;
ods graphics off;
```

Appendix C. SAS Code for Logistic Regression and Survival Analysis 2006-2015 Data Broken in Bins

```
*Survival on YOS 0-10 and YOS 20 on;
DATA "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
    SET "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
        If YOS_EFY = '' then YOSBIN = 0;
        If YOS_EFY = '0' then YOSBIN = 0;
        If YOS_EFY = '1' then YOSBIN = 0;
        If YOS_EFY = '2' then YOSBIN = 0;
        If YOS_EFY = '3' then YOSBIN = 0;
        If YOS_EFY = '4' then YOSBIN = 0;
        If YOS_EFY = '5' then YOSBIN = 0;
        If YOS_EFY = '6' then YOSBIN = 0;
        If YOS_EFY = '7' then YOSBIN = 0;
        If YOS_EFY = '8' then YOSBIN = 0;
        If YOS_EFY = '9' then YOSBIN = 0;
        If YOS_EFY = '10' then YOSBIN = 0;
        If YOS_EFY = '11' then YOSBIN = 0;
        If YOS_EFY = '12' then YOSBIN = 0;
        If YOS_EFY = '13' then YOSBIN = 0;
        If YOS_EFY = '14' then YOSBIN = 0;
        If YOS_EFY = '15' then YOSBIN = 0;
        If YOS_EFY = '16' then YOSBIN = 0;
        If YOS_EFY = '17' then YOSBIN = 0;
        If YOS_EFY = '18' then YOSBIN = 0;
        If YOS_EFY = '19' then YOSBIN = 0;
        If YOS_EFY = '20' then YOSBIN = 0;
        If YOS_EFY = '21' then YOSBIN = 1;
        If YOS_EFY = '22' then YOSBIN = 1;
        If YOS_EFY = '23' then YOSBIN = 1;
        If YOS_EFY = '24' then YOSBIN = 1;
        If YOS_EFY = '25' then YOSBIN = 1;
        If YOS_EFY = '26' then YOSBIN = 1;
        If YOS_EFY = '27' then YOSBIN = 1;
        If YOS_EFY = '28' then YOSBIN = 1;
        If YOS_EFY = '29' then YOSBIN = 1;
        If YOS_EFY = '30' then YOSBIN = 1;
        If YOS_EFY = '31' then YOSBIN = 1;
        If YOS_EFY = '32' then YOSBIN = 1;
        If YOS_EFY = '33' then YOSBIN = 1;
        If YOS_EFY = '34' then YOSBIN = 1;
        If YOS_EFY = '35' then YOSBIN = 1;
        If YOS_EFY = '36' then YOSBIN = 1;
        If YOS_EFY = '37' then YOSBIN = 1;
        If YOS_EFY = '38' then YOSBIN = 1;
        If YOS_EFY = '39' then YOSBIN = 1;
        If YOS_EFY = '40' then YOSBIN = 1;
RUN;
*Survival function;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput2.doc";
```

```
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"(where=(retain=0)
) plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;

*Divide data set into 2YOSBIN;
data "I:\setup\Desktop\Thesis\SASCombine\YOS0_2BINS";
set "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      if YOSBIN=0;
run;
data "I:\setup\Desktop\Thesis\SASCombine\YOS1_2BINS";
set "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      if YOSBIN=1;
run;

*Logistic Regression Data in YOSBIN0_2BINS;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputYOSBIN0_2BIN_LogReg.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\YOS0_2BINS";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;

*Logistic Regression Data in YOSBIN1_2BINS;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputYOSBIN1_2BIN_LogReg.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\YOS1_2BINS";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Survival function YOSBIN0_2Bin;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputSurvivalYOSBIN0_BIN.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\YOS0_2BINS"(where=(retain=0))
plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
```

```
run;
ODS RTF CLOSE;
ods graphics off;


*Survival function YOSBIN1_2Bin;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputSurvivalYOSBIN1_BIN.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\YOS1_2BINS"(where=(retain=0))
plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;


*Cox Regression Model YOSBIN0_2BIN;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputCoxReg_YOSBIN0_2BIN.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\YOS0_2BINS"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;

*Cox Regression Model YOSBIN0_2BIN;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputCoxReg_YOSBIN1_2BIN.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\YOS1_2BINS"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;


*Survival on YOS 0-4, 5-10, 11-20 and YOS 21+;
DATA "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
    SET "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
        If YOS_EFY = '' then YOSBIN1 = 0;
        If YOS_EFY = '0' then YOSBIN1 = 0;
        If YOS_EFY = '1' then YOSBIN1 = 0;
        If YOS_EFY = '2' then YOSBIN1 = 0;
```

```
        If YOS_EFY = '3' then YOSBIN1 = 0;
        If YOS_EFY = '4' then YOSBIN1 = 0;
        If YOS_EFY = '5' then YOSBIN1 = 1;
        If YOS_EFY = '6' then YOSBIN1 = 1;
        If YOS_EFY = '7' then YOSBIN1 = 1;
        If YOS_EFY = '8' then YOSBIN1 = 1;
        If YOS_EFY = '9' then YOSBIN1 = 1;
        If YOS_EFY = '10' then YOSBIN1 = 1;
        If YOS_EFY = '11' then YOSBIN1 = 2;
        If YOS_EFY = '12' then YOSBIN1 = 2;
        If YOS_EFY = '13' then YOSBIN1 = 2;
        If YOS_EFY = '14' then YOSBIN1 = 2;
        If YOS_EFY = '15' then YOSBIN1 = 2;
        If YOS_EFY = '16' then YOSBIN1 = 2;
        If YOS_EFY = '17' then YOSBIN1 = 2;
        If YOS_EFY = '18' then YOSBIN1 = 2;
        If YOS_EFY = '19' then YOSBIN1 = 2;
        If YOS_EFY = '20' then YOSBIN1 = 2;
        If YOS_EFY = '21' then YOSBIN1 = 3;
        If YOS_EFY = '22' then YOSBIN1 = 3;
        If YOS_EFY = '23' then YOSBIN1 = 3;
        If YOS_EFY = '24' then YOSBIN1 = 3;
        If YOS_EFY = '25' then YOSBIN1 = 3;
        If YOS_EFY = '26' then YOSBIN1 = 3;
        If YOS_EFY = '27' then YOSBIN1 = 3;
        If YOS_EFY = '28' then YOSBIN1 = 3;
        If YOS_EFY = '29' then YOSBIN1 = 3;
        If YOS_EFY = '30' then YOSBIN1 = 3;
        If YOS_EFY = '31' then YOSBIN1 = 3;
        If YOS_EFY = '32' then YOSBIN1 = 3;
        If YOS_EFY = '33' then YOSBIN1 = 3;
        If YOS_EFY = '34' then YOSBIN1 = 3;
        If YOS_EFY = '35' then YOSBIN1 = 3;
        If YOS_EFY = '36' then YOSBIN1 = 3;
        If YOS_EFY = '37' then YOSBIN1 = 3;
        If YOS_EFY = '38' then YOSBIN1 = 3;
        If YOS_EFY = '39' then YOSBIN1 = 3;
        If YOS_EFY = '40' then YOSBIN1 = 3;
RUN;
*Survival function;
ods graphics on;
ODS RTF FILE="I:\setup\Desktop\Thesis\SASOutput3.doc";
axis2 order = (1 to 40 by 1);
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL"(where=(retain=0)
) plots=survival(atrisk);
by CTPR
by NOTSORTED YOSBIN1;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;

*Divide data set into 4YOSBIN;
```

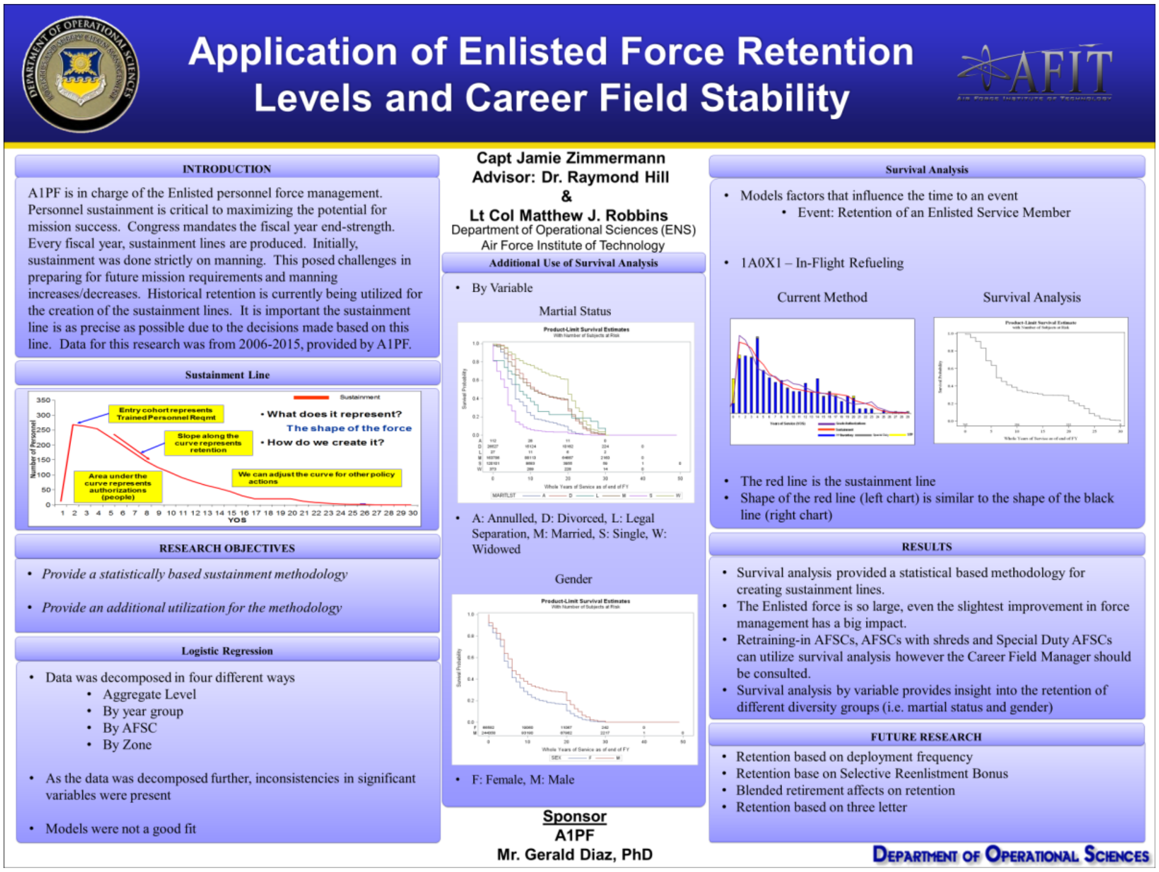
```
data "I:\setup\Desktop\Thesis\SASCombine\YOS0_4BINS";
set "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      if YOSBIN1=0;
run;
data "I:\setup\Desktop\Thesis\SASCombine\YOS1_4BINS";
set "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      if YOSBIN1=1;
run;
data "I:\setup\Desktop\Thesis\SASCombine\YOS2_4BINS";
set "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      if YOSBIN1=2;
run;
data "I:\setup\Desktop\Thesis\SASCombine\YOS3_4BINS";
set "I:\setup\Desktop\Thesis\SASCombine\ALLDATAFINAL";
      if YOSBIN1=3;
run;



*Logistic Regression Data in YOSBIN0_4BINS;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputYOSBIN0_4BIN_LogReg.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\YOS0_4BINS";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;

*Logistic Regression Data in YOSBIN1_4BINS;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputYOSBIN1_4BIN_LogReg.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\YOS1_4BINS";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Logistic Regression Data in YOSBIN2_4BINS;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputYOSBIN2_4BIN_LogReg.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\YOS2_4BINS";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;


*Logistic Regression Data in YOSBIN3_4BINS;
```

```
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputYOSBIN3_4BIN_LogReg.doc";
proc logistic data = "I:\setup\Desktop\Thesis\SASCombine\YOS3_4BINS";
      class SEX GRADE RACE MARITLST DEPENDS;
      model retain = GRADE SEX RACE MARITLST DEPENDS YOS_EFY;
run;
ODS RTF CLOSE;
ods graphics off;



*Survival function YOSBIN0_4Bin;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputSurvivalYOSBIN0_4BIN.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\YOS0_4BINS"(where=(retain=0))
plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;


*Survival function YOSBIN1_4Bin;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputSurvivalYOSBIN1_4BIN.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\YOS1_4BINS"(where=(retain=0))
plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;

*Survival function YOSBIN2_4Bin;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputSurvivalYOSBIN2_4BIN.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\YOS2_4BINS"(where=(retain=0))
plots=survival(atrisk);
axis2 order = (1 to 40 by 1);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
```

```
ODS RTF CLOSE;
ods graphics off;


*Survival function YOSBIN3_4Bin;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputSurvivalYOSBIN3_4BIN.doc";
proc lifetest
data="I:\setup\Desktop\Thesis\SASCombine\YOS3_4BINS"(where=(retain=0))
plots=survival(atrisk);
by CTPR;
title 'Survival Graph for #byval(CTPR)';
time YOS_EFY*retain(1);
run;
ODS RTF CLOSE;
ods graphics off;



*Cox Regression Model YOSBIN0_4BIN;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputCoxReg_YOSBIN0_4BIN.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\YOS0_4BINS"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;


*Cox Regression Model YOSBIN1_4BIN;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputCoxReg_YOSBIN1_4BIN.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\YOS1_4BINS"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;

*Cox Regression Model YOSBIN2_4BIN;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputCoxReg_YOSBIN2_4BIN.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\YOS2_4BINS"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
```

```
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;


*Cox Regression Model YOSBIN3_4BIN;
ods graphics on;
ODS RTF
FILE="I:\setup\Desktop\Thesis\SASOutputCoxReg_YOSBIN3_4BIN.doc";
proc phreg data = "I:\setup\Desktop\Thesis\SASCombine\YOS3_4BINS"
plots=survival;
class SEX GRADE RACE MARITLST DEPENDS;
model YOS_EFY*retain(0) = SEX GRADE RACE MARITLST DEPENDS;
by CTPR;
run;
ODS RTF CLOSE;
ods graphics off;
```

Appendix D. Storyboard



Bibliography

Brauner, M., Lawson, K., Mickleson, W., Adam, J. & Chaiken. (1991). "Time series models for predicting monthly losses of Air Force enlisted personnel." Rand Corporation, Santa Monica, California

Cashbaugh, D. (2010). Managing Retention Use of Simulation and Optimization. Briefing slides retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a535605.pdf.

Draper, N. R. & Smith, H. (2014). Applied regression analysis. John Wiley & Sons.

Jantcher, H. 2016. An examination of economic metrics as Indicators of Air Force retention. MS Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH.

Jonas, R. (1979). "The Personnel Management Model". Retrieved from http://www.dtic.mil/get-tr-doc/pdf?AD=ADA077038.

French, S. (2000). Analyzing Personnel Retention Utilizing Multi-Agent Systems. Retrieved from http://www.dtic.mil/docs/citations/ADA386919.

Gildea, D. (2013). "Air Force Personnel Center: MilPDS Upgrade Begins". Retrieve from http://www.af.mil/News/ArticleDisplay/tabid/223/Article/109624/milpds-upgrade-begins.aspx.

Hill, R.R. and Gaupp, M.P. 2006. Visualization Tools for Perspective Analysis of a Complex Adaptive System Model of Pilot Retention. *International Journal of Modeling and Simulation, 6*(4):378-384.

Montgomery, D., Peck, E. & Vining G. (2012). Introduction to Linear Regression Analysis, 5th edition. John Wiley & sons Inc. Hobkon, New Jersey.

Orrick, S. (2008). Forecasting Marine Corps Enlisted Losses. Naval Post Graduate School, Monterey California.

Sainani, K. (n.d). Introduction to Cox Regression. Briefing slides retrieved from www.pitt.edu/~super4/33011-34001/33051-33061.ppt.

Schofield, J. (2015). “Non-Rated Air Force Line Officer Attrition Rates Using Survival Analysis”. Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio.

Train, K. E. (2002). Discrete choice methods with simulation. Cambridge university press.

Whelan, G. (2013). Forecasting Army Enlisted ETS Losses. Naval Postgraduate School. Monterey, California.

Wiggins V., Engquist, S. & Looper, T. (1992). Applying Neural Networks to Air Force Personnel Analysis. Retrieved from http://www.dtic.mil/docs/citations/ADA248762.

Zens, C. (2016). “Application of Non-Rated Line Officer Attrition Levels and Career Field Stability”. Air Force Institute of Technology, Wright Patterson Air Force Base, Ohio.

REPORT DOCUMENTATION PAGE							*Form Approved* *OMB No. 074-0188*
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**							
1. REPORT DATE *(DD-MM-YYYY)* 23-03-2017			**2. REPORT TYPE** Master's Thesis				**3. DATES COVERED** *(From – To)* September 2015 – March 2017
TITLE AND SUBTITLE Application of Enlisted Force Retention Levels and Career Field Stability							**5a. CONTRACT NUMBER**
							5b. GRANT NUMBER
							5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Zimmermann, Jamie T., Capt, USAF							**5d. PROJECT NUMBER**
							5e. TASK NUMBER
							5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865							**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENS-MS-17-M-167
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AGENCY: Force Management & Enterprise Readiness Analysis Div ADDRESS: Andrews AFB, MD PHONE and EMAIL: (240) 612 – 4040 ATTN: Dr. Gerald Diaz, PhD							**10. SPONSOR/MONITOR'S ACRONYM(S)** HAF/A1PF
							11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **DISTRUBTION STATEMENT A.** APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.							
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.							
14. ABSTRACT The Air Force's success is influenced by the manpower it has. The end-strength is Congressionally mandated. Over 10 years ago, the Air force produced enlisted career field sustainment lines based on manning. If a career field was over manned the authorizations were decreased, likewise if a career field was under manned the authorizations were increased. The constant fluctuation of manning caused bathtubs to be created and requirements to go unfilled. Currently, the Air force produces enlisted career field sustainment lines based on the 5-year historical retention rates. This method produced a more steady state approach, as well as providing a means to adjust the line for other policy actions such as retraining in/out. The need to have a statistically base approach is essential for explaining and defending the creation of the sustainment lines. Data from 2006-2015 was utilized in this research. Logistic regression was used to determine if any significant variables existed, however logistic regression did not provide insight into the behavior of the data. A survival analysis approach, using retention data, provides a statistically sound methodology to the creation of the sustainment lines. This study produces sustainment lines based on the survival functions for each enlisted career field.							
15. SUBJECT TERMS enlisted force, retention, logistic regression, survival analysis							
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr. Raymond Hill, Jr. AFIT/ENS		
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	73	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-6565 x7469 Raymond.Hill@afit.edu		

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18